Mobile small RNAs regulate genome-wide DNA methylation

BIOLOGICAL SCIENCES: Plant Biology

Mobile small RNAs regulate genome-wide DNA methylation

Mathew G. Lewsey[1,2,3], Thomas J. Hardcastle[3,4], Charles W. Melnyk[4,5], Attila Molnar[4,6], Adrián Valli[4], Mark A. Urich[1], Joseph R. Nery[1], David C. Baulcombe[4,7*] and Joseph R. Ecker[1,2,7,8*]

[1]Genomic Analysis Laboratory, The Salk Institute for Biological Studies, La Jolla, California 92037, USA
[2]Plant Biology Laboratory, The Salk Institute for Biological Studies, La Jolla, California 92037, USA
[3]Co-first author
[4]Department of Plant Sciences, University of Cambridge, Downing Street, Cambridge CB2 3EA, UK
[5]Present address: Sainsbury Laboratory, University of Cambridge, Bateman Street, Cambridge CB2 1LR, UK
[6]Present address: School of Biological Sciences, The University of Edinburgh, Institute of Molecular Plant Sciences, Mayfield Road, Edinburgh EH9 3JR, UK
[7]Co-corresponding author
[8]Howard Hughes Medical Institute, The Salk Institute for Biological Studies, La Jolla, California 92037, USA

*Correspondence: ecker@salk.edu (J.R.E.), dcb40@cam.ac.uk (D.C.B.)

Keywords: RNA-directed DNA methylation, plant grafting, transposable element, small RNA, transcriptional gene silencing

Abstract

RNA silencing at the transcriptional and post-transcriptional levels regulates endogenous gene expression, controls invading transposable elements (TEs) and protects the cell against viruses. Key components of the mechanism are small RNAs (sRNAs) of 21-24 nucleotides (nt) that guide the silencing machinery to their nucleic acid targets in a nucleotide sequence-specific manner. Transcriptional gene silencing is associated with 24 nt sRNAs and RNA-directed DNA methylation (RdDM) at cytosine residues in three DNA sequence contexts (CG, CHG and CHH). We previously demonstrated that 24 nt sRNAs are mobile from shoot to root in *Arabidopsis thaliana* and confirmed that they mediate DNA methylation at three sites in recipient cells. In this study we extend that finding by demonstrating that RdDM of thousands of loci in root tissues is dependent upon mobile sRNAs from the shoot and that mobile sRNA-dependent DNA methylation occurs predominantly in non-CG contexts. Mobile sRNA-dependent non-CG methylation is largely dependent on the DRM1/DRM2 RdDM pathway but independent of the CMT2/3 DNA methyltransferases. Specific superfamilies of TEs, including those typically found in gene-rich euchromatic regions, lose DNA methylation in a mutant lacking 22 to 24 nt sRNAs (*dicer-like 2, 3, 4* triple mutant). Transcriptome analyses identified a small number of genes whose expression in roots is associated with mobile sRNAs and connected to DNA methylation directly or indirectly. Finally, we demonstrate that sRNAs from shoots of one accession move across a graft union and target DNA methylation *de novo* at normally unmethylated sites in the genomes of root cells from a different accession.

Significance statement

Small RNAs (sRNAs) of 24 nucleotides are associated with transcriptional gene silencing by targeting DNA methylation to complementary sequences. We demonstrated previously sRNAs move from shoot to root, where they regulate DNA methylation of three endogenous transposable elements (TEs). However, the full extent of root DNA methylation dependent on mobile sRNAs was unknown. We demonstrate that DNA methylation at thousands of sites depends upon mobile sRNAs. These sites are associated with TE superfamilies found in gene rich regions of the genome, which lose methylation selectively in an sRNA-deficient mutant. If the TEs were able to reactivate they could cause genome instability and altered gene expression patterns, with negative effects on the plant. Consequently, mobile sRNAs may defend against these TEs.

\body

Introduction

RNA silencing in plants and animals is a process that controls gene expression at both the transcriptional and post-transcriptional levels (1). In plants small non-coding RNAs (sRNAs), 21 to 24 nucleotides (nt) in length (2, 3), direct the RNA silencing machinery to target nucleic acids in a sequence-specific manner (4). The 21/22 nt sRNAs are primarily associated with mRNA cleavage and are involved in post-transcriptional gene silencing (PTGS) (3). The 24 nt sRNAs are primarily associated with RNA-directed DNA methylation (RdDM) and transcriptional gene silencing (2, 5). However, a recent study suggests that both 21 and 24 nt sRNAs are involved in deposition of DNA methylation (6). It is proposed that the 21 nt sRNAs may establish DNA methylation, whereas the 24 nt species are involved in its amplification and maintenance (7, 8).

RdDM involves methylation of cytosine residues in CG, CHG and CHH sequence contexts (where H denotes any base except G) (9, 10). It is closely associated with repressive chromatin marks at target loci (11) and blocks gene transcription when present in promoter regions (12). RdDM maintains genome integrity by repression of transposable element (TE) activity, as well as contributing to environmental and developmental regulation of gene expression (4, 8, 13-17). The methylation status of DNA is heritable through both meiosis and mitosis, allowing it to persistently alter gene expression (18-20).

Initial establishment of RdDM involves cleavage of double-stranded RNA by DICER-LIKE (DCL) proteins to form 21 nt to 24 nt sRNAs, which load into ARGONAUTE (AGO) proteins (4, 21). These nucleoprotein complexes target chromatin-associated scaffold transcripts in a sequence-specific manner (22-24). The chromatin-bound complexes then recruit DOMAINS REARRANGED METHYLTRANSFERASES 1 and 2 (DRM1 and DRM2) that methylate DNA in CG, CHG and CHH sequence contexts (8, 25). A complex set of maintenance mechanisms ensures persistence of established DNA methylation through cell division and even between generations. Most of these mechanisms are independent of RNA and involve epigenetic histone marks. The VARIANT IN METHYLATION (VIM) family proteins 1, 2 and 3 and DNA METHYLTRANSFERASE 1 (MET1) maintain efficiently CG context methylation, resulting in near complete methylation of target sequences (26-28). Non-CG context methylation (i.e. CHG, CHH) is maintained by a self-reinforcing loop involving KRYPTONITE family enzymes (29). These proteins recognize non-CG context methylated DNA and methylate lysine 9 of adjacent histone H3 (H3K9me2). CHROMOMETHYLASE (CMT) proteins CMT2 and CMT3 bind H3K9me2 and methylate adjacent non-CG sites (30) of the newly replicated DNA. Redundancy exists between target sites of CMT2 and CMT3, but their predominant functions are to maintain CHH and CHG context methylation respectively (31). The activity of CMT2 is substantially less efficient than that of MET1 so that CMT2 target sites exhibit variable levels of DNA methylation. CMT3 efficiency is intermediate between MET1 and CMT3. Both CMT2 and CMT3 typically target long TEs and gene distal TEs (31).

There are also RNA-dependent mechanisms to maintain DNA methylation that involve plant-specific DNA-dependent RNA polymerases IV and V (POL IV and POL V). These polymerases are recruited to chromatin by methyl-DNA-binding proteins SHH1 and SUVH2 and 9 (32). POL IV produces precursor RNAs that are processed into 24 nt sRNAs, whereas POL V produces chromatin bound scaffold transcripts at sites of DNA methylation (33). Together, they ensure maintenance of CG, CHG and CHH context DNA methylation through an AGO-dependent mechanism similar to the mechanism that establishes methylation, in which an AGO-sRNA recruits DRM1 and DRM2 to maintain non-CG DNA methylation (8, 25). The target sites of the DRM1/DRM2 DNA methylation maintenance pathway are largely non-overlapping with those of CMT2 and CMT3, and tend to be short, gene proximal TEs and the edges of long TEs (31, 34).

RdDM can operate cell-to-cell and systemically due to translocation of 23-24 nt sRNAs from shoots to roots (35). In our previous studies we confirmed that these mobile 23-24 nt sRNAs

target RdDM and TGS at one transgene and RdDM at three endogenous TEs (35-37). Depletion of shoot sRNAs corresponded with reduction of 23-24 nt sRNAs in wild-type roots, indicating shoot-derived sRNAs contribute to the total root sRNA population (36). In this study we investigate the extent to which mobile sRNAs mediate genome-wide RdDM. Our approach, as before, is to analyze sRNA and DNA methylation in roots of grafted plants that are defective for the production of the 24 nt sRNA species associated with RdDM. The sRNAs in these grafted plants move predominantly from shoot to root following source-sink gradients (36) and, by grafting different genotypes as shoots, we identify changes in DNA methylation and gene expression in the roots that are dependent on mobile sRNAs.

We show that mobile sRNAs influence genomic DNA methylation at thousands of loci, and that the affected loci are predominantly associated with transposons of specific classes. A very small number of protein coding genes were influenced by this mobile RdDM. The mobile sRNA-dependent DNA methylation is associated with the DRM1/DRM2 RdDM pathway but not the CMT2/3 DNA methyltransferase pathway. Furthermore, we demonstrate that mobile sRNAs unique to one accession established DNA methylation *de novo* in unmethylated regions of the genome of a second accession.

Results

Identification of DNA methylation loci targeted by mobile sRNAs

The primary aims of our study were to determine how many genomic loci in *Arabidopsis* may be targeted by RdDM due to the direct action of mobile sRNAs (Fig. 1) and to compare these loci with other sRNA targeted and cytosine methylated regions throughout the genome. We reasoned that mobile sRNAs that direct RdDM may be associated with specific features of the genome and depend upon specific genes in the RdDM pathway. In addition, we aimed to determine how many of these directly-affected loci might influence gene expression. It could be, for example, that DNA methylation, leading to TGS, of a transposon target of mobile sRNA affects expression of an adjacent gene.

In previous work we compared sRNA populations in roots of WT *Arabidopsis thaliana* (accessions Col-0, termed Col, and C24) and a *dicer-like 2 dicer-like 3 dicer-like 4* triple mutant (Col background, termed *dcl234*) that had been grafted to C24, Col and *dcl234* shoots in order to identify loci producing mobile sRNAs in the *Arabidopsis* genome. The *dcl234* triple mutant is unable to produce 22 to 24 nt sRNAs associated with RNA silencing (23). We reasoned that any 22 to 24 nt sRNAs present in *dcl234* roots grafted to WT shoots must have moved from the grafted WT shoot (36). Loci similarly represented by sRNA reads in C24/C24, C24/Col, Col/Col and C24/*dcl234* but not *dcl234*/*dcl234* datasets (notation shoot/root, with underline indicating the analyzed plant part) could be confidently assigned as producing mobile sRNA, since the WT shoot complements the inability of *dcl234* roots to produce sRNA. Contrastingly, the sRNAs from loci that were absent in C24/*dcl234* and *dcl234*/*dcl234* grafts and present in grafts with Col or C24 roots were interpreted as being dependent on *DCL2,3,4* but not mobile. Loci where sRNAs were present in all datasets indicate sRNAs produced by DCL1 or in a DCL-independent manner.

We incorporated genome-wide MethylC-seq data in our analyses and applied the above reasoning to classify the genome-wide patterns of DNA methylation affected by graft transmissible sRNAs. We used five genotype combinations in two-way grafts to define six locus classes (A-F) by their combinations of DNA methylation and sRNA representation across graft combinations (Fig. 1, where high and low sRNA abundance are indicated by dark or light blue, respectively, and high and low DNA methylation levels by dark or light red, respectively; also SI Appendix Figs. S1, S2). Three classes of these were of primary interest:

A. "Direct" loci correspond to overlapping regions of cytosine methylation and sRNAs present in C24/C24, C24/Col, Col/Col and C24/*dcl234* but not *dcl234*/*dcl234*. This model is consistent with direct targeting of DNA methylation by mobile sRNAs.
B. "Indirect" loci show the same pattern of DNA methylation between grafts as "Direct" loci but these were not associated with sRNAs. Here we infer that there are undetectable levels of mobile sRNAs or that the change in methylation is due to indirect effects of a mobile signal.
C. "*De novo* methylated" loci correspond to overlapping regions of sRNAs and cytosine methylation present in C24/C24, C24/Col and C24/*dcl234*, but not in *dcl234*/*dcl234* or Col/Col. These loci likely correspond to sRNAs produced uniquely by C24 that target DNA methylation to previously unmethylated regions of the Col genome.

The class A (direct) and C (*de novo*) loci because these are most likely to be associated with the mobile sRNA. We compared their characteristics with those of class B (indirect) loci that are associated with a mobile signal but, by definition, do not correspond with detectable candidate sRNAs. Three classes of loci (D-F) not associated with regulation of DNA methylation by mobile sRNA were also identified and are described in the Supplemental Results.

Mobile sRNA and DNA methylation

We developed a statistical framework that could be applied to sRNA and MethylC-seq data to accurately assess the models described in Fig. 1 (see Supplemental Materials and Methods sections 1, 3 and 4 for detailed explanation). Our previous sRNA abundance data (36) were reanalyzed using this framework to ensure consistency with analyses of the MethylC-seq datasets generated in the current study (Table 1). DNA methylation associated with mobile sRNA was examined separately by sequence context (CG, CHG, CHH), because the genetic requirements for each are distinct.

Our analyses revealed that non-CG DNA methylation is abundant at class A (direct) loci; we identified 13 CG, 398 CHG and 401 CHH-context class A loci with a false discovery rate (FDR) of 5% (Table 1, Dataset S1). These included the three loci we identified previously as possessing mobile sRNA associated DNA methylation (36). The association between regions of sRNA and DNA methylation that corresponded to the class A loci model was statistically significantly more frequent than expected by chance (Table 1, Z-scores; CG 154.69, CHG 77.123, CHH 84.250; $p<0.05$). Furthermore, class A loci of all three DNA methylation contexts overlapped statistically significantly (SI Appendix Fig. S3). Observation of co-localized DNA methylation in all three contexts is characteristic of RdDM. Fewer than 2.5% of the class A loci overlapped with regions identified as variable over 30 generations in *Arabidopsis* (18) and thus it is unlikely that the results are due to spontaneous epiallelic variation.

We investigated the genes potentially regulating DNA methylation at class A loci by analysis of MethylC-seq data from 86 *Arabidopsis* gene silencing mutant lines (31). We examined the effects of mutation of a subset of genes with known roles in RdDM (*dcl2/3/4*, *met1*, *suvh4/5*, *ddm1, suvh4/5/6, vim1/2/3, cmt2, cmt3*, and WT control) on DNA methylation at class A loci. We focused on non-CG methylation patterns because very few CG-context class A loci were identified. Mutation of *drm1/2* caused the strongest reduction in non-CG context methylation in class A loci (Fig. 2, SI Appendix Fig. S4). The other mutations examined showed weaker effects, of which *cmt2* and *cmt3* caused the least changes in DNA methylation; *cmt3* reduced CHG context DNA methylation to levels intermediate between WT and *drm1*/*drm2* but had no effect in the CHH context, whilst *cmt2* had no effect in either context. DNA methylation was reduced to levels intermediate between WT and *drm1/2* in both sequence contexts by all others of these mutations. It should be noted that the MethylC-seq data from gene silencing mutant lines were generated using leaves of three-week old plants, whereas our MethylC-seq data were generated using roots of three-week old plants. We considered it appropriate to compare the two datasets because there is little evidence for tissue-specific variation in DNA methylation in *Arabidopsis*. We

examined the similarity of DNA methylation in these two tissues by making comparisons between the likelihoods of methylation at loci identified as methylated in root tissue of the Col/Col graft from our dataset and in the three-week old Col-0 (WT) leaf tissue of Stroud and colleagues (31). DNA methylation was broadly consistent between the two tissues (Dataset S2, SI Appendix Fig. S5). Additionally, a prior study demonstrated that DNA methylomes from two tissues in each of eleven *Arabidopsis* accessions cluster by accession, not tissue type (38).

DNA methylation under indirect regulation by the mobile signal

Class B ("indirect") loci show the same pattern of DNA methylation between grafts as do class A ("direct") loci. However, they are distinguished by the lack of associated sRNA. We found that class B and class A loci had extremely similar characteristics. Hundreds of class B loci were detected almost exclusively in non-CG contexts (13 CG, 868 CHG, 1029 CHH; Table 1, Dataset S1). These loci were unambiguously identified and located in the genome. We found fewer than 0.5% of the class B loci overlapped with regions previously identified as spontaneously variable between parent and offspring plants, suggesting our results are not a consequence of this phenomenon (18).

The distribution of the class B loci across DNA methylation contexts mirrors that of the class A loci (13 CG, 398 CHG and 401 CHH), though substantially more class B loci were identified. We focused on non-CG class B loci since so few CG loci were identified. The association between regions lacking sRNAs and possessing DNA methylation, corresponding to the class B loci model, was lower than would be expected by chance (Table 1; CHG, p=0, Z=-92.603; CHH, p=0, Z=-105.48). This pattern suggests that methylation loci regulated by a mobile signal are negatively associated with regions lacking in small RNAs.

The class A and B loci were also influenced similarly by mutations in RdDM pathway genes (31) (Fig. 2, SI Appendix Fig. S4). Mutation of *drm1/2* caused the strongest reduction in non-CG context DNA methylation (Fig. 2, SI Appendix Fig. S4). All other mutations examined reduced non-CG context methylation at class B loci except *cmt2*. Notably, *cmt3* had the next weakest effect, reducing CHG context DNA methylation but not affecting CHH context methylation.

In summary, there was very little to distinguish class A and class B loci, suggesting they are ultimately under the regulation of the same mobile signal. In principle this signal could be an as yet unidentified secondary regulatory factor dependent on the DCL234-dependent sRNA at the class A loci. Alternatively it could be caused by the very low levels of DCL234-dependent sRNA that fail to pass the significance thresholds in our model (SI Appendix Figs. S6-8), or which are not well sequenced.

We used an independent statistical approach to estimate the total number of class A and B loci in the dataset, as a complement to the results above (see Supplemental Materials and Methods section 5 for detailed explanation). This approach evaluates all loci that may exist without specifying genomic coordinates. We estimate that there exist 72 CG (59-85 with 95% confidence), 1557 CHG (1506-1608 with 95% confidence) and 2238 CHH (2170-2303 with 95% confidence) loci of class A, and 526 CG (485-568 with 95% confidence), 5699 CHG (5597-5802 with 95% confidence) and 10943 CHH (10785-11099 with 95% confidence) loci of class B by this approach. From these data we conclude that thousands of DNA methylation loci found predominantly in the non-CG context may be regulated by the mobile signal.

DNA methylation regulated by the mobile signal associates with specific features of the genome

Non-CG context class A loci were significantly associated with TEs and promoters containing TEs, but not with promoters that do not contain TEs (Fig. 3A). Contrastingly, they were significantly depleted in genes, coding regions and 5' UTRs. CG context class A loci were significantly associated with TEs only. Further analysis showed that class A loci associated

significantly with many superfamilies of TE (Fig. 3B). RAths, LINE and SINE superfamily TEs were targeted more strongly than other TEs, most clearly so by the non-CG context class A loci. TEs are broadly grouped according to their replication strategies as Type 1 (retroelements) or Type 2 (DNA elements), and the targeted superfamilies are all short, gene-proximal Type 1 TEs. Class A loci were targeted most highly by 23-24 nt mobile sRNAs, eliminated in *dcl234*/*dcl234* grafts (Fig. 3C). These results indicate that DNA methylation directly targeted by the shoot-root mobile sRNA is associated with specific TE superfamilies.

Non-CG class B loci were also associated significantly with TEs and promoters containing TEs (Fig. 4A, SI Appendix Fig S2), as were non-CG class A loci. They were associated with many different TE superfamilies, most strongly so with certain Type 1 retroelement superfamilies (RAths, LINE, SINE). They also showed clear associations with specific retroelement superfamilies (Type 2 TEs) including DNA Mariner, Pogo and HAT.

The genome features associated with class A and B loci might either be specific to these locus classes and to the mobile signal or they could be driven by dependence of the loci upon DCL234-derived sRNAs. To address this point we compared the features associated with class A and B loci to all features that lose DNA methylation in *dcl234*/*dcl234* grafts. The features that lost DNA methylation in *dcl234*/*dcl234* grafts were highly similar to those associated with class A and B loci (SI Appendix Fig. S9). The data indicate a specific subset of TEs lose DNA methylation in the *dcl234* mutant. Furthermore, they suggest that dependency upon DCL234 drives the features associated with class A and B loci rather than these features being specific to the mobile signal.

The influence of the mobile signal on gene expression

We examined whether DNA methylation regulated by the mobile signal may influence gene expression. Root transcriptomes of the graft combinations described by Fig. 1 and SI Appendix Fig S1 (excluding C24/C24) were profiled by RNA-seq. A total of 23 transcripts were identified as significantly differentially regulated in a manner that indicated association with the mobile signal (Datasets S3A, S3B). These included transcripts coding for proteins involved in cell wall formation, membrane transport and regulation of osmotic potential. Three of the genes encoding these were associated with DNA methylation regulated either directly or indirectly by the mobile signal; AT2G01880 - PURPLE ACID PHOSPHATASE 7, AT2G36490 - REPRESSOR OF SILENCING 1 (*ROS1*), AT3G43270 – pectin methylesterase. We measured the number of differentially expressed genes between C24/Col and *dcl234*/*dcl234* roots as a control. Twenty-seven transcripts were identified as significantly differentially regulated in roots of *dcl234* compared to WT roots (Datasets S4A, S4B). Differential expression of *ROS1* (AT2G36490.1) was confirmed in roots of *dcl234*/*dcl234* *versus* C24/*dcl234* and C24/Col by quantitative real time PCR (Q-RT-PCR, SI Appendix Fig. S10). These findings demonstrate that loss of *DCL2*, *DCL3* and *DCL4* causes differential expression of a small number of genes, consistent with data in previous reports (39).

We investigated whether transcripts dependent upon *DCL234* (Datasets S4A, S4B) or associated with the mobile signal (Datasets S3A, S3B) might also depend upon *DRM1/2.* This was done by identifying transcripts differentially regulated between WT and the *drm1/2* mutant using published RNA-seq data. These data were generated from leaves of three-week old plants in parallel with the MethylC-seq data discussed by Fig. 3 and SI Appendix Fig. S4 (31). Eighty-six transcripts were significantly differentially regulated between WT and *drm1/2* plants (Dataset S5). The transcript of *ROS1* (AT2G36490) was shared between these and transcripts associated with the mobile signal (Dataset S3B). The transcript AT1G53480.1 was dependent upon *DRM1/2* (Dataset S5) and *DCL234* (Dataset S4A). We attribute the low degree of overlap between *DRM1/2*-dependent and mobile signal associated transcripts to tissue specificity of gene expression. Our sRNA-seq and MethylC-seq experiments were conducted on roots of three-week old plants. As discussed above, comparison between DNA methylomes from different tissues is appropriate because there is little evidence for tissue specific variation in DNA methylation. However, transcript abundance differs more between tissues. This was demonstrated by comparison of

RNA-seq data from two tissues of multiple *Arabidopsis* accessions, where transcript abundance correlated more highly between like tissues of different accessions than different tissues of one accession (38).

Epialleles are transmitted across accessions by mobile sRNAs

There is natural variation in RdDM and its associated sRNAs in *Arabidopsis* and the term epiallele describes genomic locations where methylation differs between accessions (38, 40, 41). We investigated whether sRNAs produced uniquely by C24, and not by Col, could direct DNA methylation *de novo* at previously unmethylated sites in the Col genome (Fig. 1; class C loci). Two class C loci were unambiguously identified in the CHG DNA methylation context and four in the CHH context (Dataset S1). Three distinct sites of *de novo* RdDM existed after taking account for overlap of these loci between DNA methylation contexts (Fig. 5). None of these loci had been previously identified as spontaneously variable between parent and offspring plants (18). These results indicate that exogenous sRNAs supplied by shoots can target *de novo* DNA methylation at unmethylated sites in the genome of root cells, thereby transmitting epiallelic states from one *Arabidopsis* accession to another.

Discussion

In this study we have characterized the genome-wide distribution of 23-24 nt shoot/root mobile sRNAs and identified regions of DNA methylation that they target directly (class A loci). We also identified regions of DNA methylation dependent indirectly upon mobile sRNAs (class B loci). We found that loss of mobile sRNAs generated in the shoots disrupts DNA methylation at thousands of sites in the roots. The class B (indirect) loci were most numerous, but the two classes were otherwise indistinguishable. Both were almost entirely in the non-CG (CHG/CHH) DNA methylation context. The class A and B loci were also significantly associated with the same TE superfamilies. We found that these TE superfamilies also lost DNA methylation in the *dcl234* mutant, which is deficient in 22 to 24 nt sRNAs, indicating that the feature association of class A and B loci was driven by the dependency of mobile sRNAs on DCL234. Our results suggest the function of mobile sRNAs is to reinforce silencing of these TEs. Furthermore, we have provided mechanistic insights into the specific RdDM pathways regulating mobile sRNA-dependent methylation. Data can be visualized in our interactive genome browser (http://neomorph.salk.edu/mobile_methylome.php).

Our data confirm that mobile sRNA-dependent DNA methylation requires *DRM1* and *DRM2*, key components of the 24 nt RdDM pathway. There was only limited dependence upon *CMT3* and *CMT2*, both of which belong to a distinct DNA methylation maintenance pathway. Furthermore, class A and B loci were associated with the shortest superfamilies of TE (Figs. 2, 4, SI Appendix Table S1) including Type 1 retroelements, such as RAth elements (E1, E2, E3), SINEs and LINEs, as well as certain Type 2 DNA elements (Mariner, Pogo, RC Helitron). The DNA methylation pathways containing *DRM1/2* and *CMT2/3* are distinct, but overlapping, and both deposit non-CG methylation (8, 34). The *DRM1/2* pathway methylates small TEs and TE edges, whereas the *CMT2/3* pathway is responsible for methylation of long TEs (31, 34). Our data are in concordance with this pattern and they establish a relationship of mobile sRNA-dependent methylation with a specific RNA silencing mechanism.

The direct and indirect (class A and B) loci are essentially identical according to our data, except that no mobile sRNAs were identified associated with indirect class B loci. These findings lead to three hypotheses for further investigation. Firstly, that the mobile sRNAs from the shoot regulate an unidentified secondary signal in the roots. This secondary signal would then directly regulate DNA methylation at the B loci. Secondly, that mobile sRNAs can direct the RNA silencing machinery to sites that they match imperfectly (i.e. with which they have mismatches in sequence homology). We required perfect matching between sRNAs and genomic sequence when identifying sRNA target loci, but it is possible that a certain degree of mismatch between mobile sRNAs and their targets is either tolerated or required in certain circumstances. However,

allowing mismatches during sRNA mapping (and therefore in targeting) permits individual sRNAs to target multiple sites due to their inherently short sequence, many of which are presumably spurious (42). Therefore, detailed analysis of individual methylation regions and thorough experimental validation is required to test this hypothesis. Moreover, it appears unlikely the conclusions of our study would be altered by the outcome of such experiments, because we find the characteristics of class A (direct) and class B (indirect) loci to be essentially identical. The third hypothesis is that these loci are targeted by mobile sRNAs of very low abundance, which could not be detected in our sRNA-seq libraries.

Class A and B loci were substantially more numerous than the other classes (Fig. 1, Table 1, SI Appendix Table S2 and Fig. S1). The class D, E and F loci described respectively loci not methylated by mobile sRNAs, not remethylated by mobile sRNAs, and loci whose methylation was independent of DCL234-derived sRNAs. Class F loci were the least numerous. This observation suggests there may be few mobile sRNA-targeted loci at which that sRNA does not regulate DNA methylation.

sRNAs of 24 nt length were proportionally the most common sRNA species detected in our sRNA-seq experiments and demonstrated the clearest shoot-root transmission (Fig. 2) (36). This study and another indicate that 21-22 nt sRNAs are also mobile and associated with DNA methylation (37). These data are consistent with previous proposals of a role for the 21-22 nt size class in RdDM (6, 7, 37, 43). Furthermore, the Type 1 retroelements that we find to be the predominant targets of mobile sRNA-dependent DNA methylation produce RNA intermediate replication stages, making it possible that if transcriptionally active they are also targets of PTGS *via* mobile 21 nt sRNAs (44). However, fewer 21-22 nt sRNAs were mobile than 24 nt sRNAs, suggesting 24 nt sRNAs are the primary shoot-root signal. The design of our experiment did not permit us to completely eliminate the possibility that “mobile” 21 nt sRNAs are produced in the roots. This is because the *dcl234* mutant contains functional DCL1, which generates 21 nt sRNAs, and the *dcl234* mutant eliminates their production only partially (23, 45-47). Detailed investigation of their contribution would involve generation of additional series of *Arabidopsis* mutant grafts and is complicated by the lethality of a *dcl1* null plant (48).

Although mobile sRNAs have minimal effects on gene expression in *Arabidopsis*, we predict that their influence would be much greater in TE rich genomes. In such genomes, including those of many common crops, mobile sRNAs may be an important mechanism of genome defense. The TE superfamilies targeted by mobile sRNAs are enriched in euchromatic regions, typically within c. 500 nt of the nearest gene (49, 50). TEs inserted near genes can have dramatic effects on gene expression (15, 50). Expression of a small number of transcripts was differentially regulated in our study, consistent with previous observations. (39). However, the *A. thaliana* genome contains substantially fewer TEs than related outcrossing species, such as *A. lyrata* (51, 52). Furthermore, TEs in the *A. thaliana* genome show a lower rate of active transposition (51). This suggests that TEs in the *A. thaliana* genome may have relatively less influence on gene expression than TEs in outcrossing species, which would result in fewer mobile sRNA-dependent regulated genes. Nonetheless, those genes that did exhibit mobile sRNA-dependent regulation had diverse functions, indicating they may have significant influence in the correct conditions. Moreover, mobile sRNAs are able to move into both meristematic and meiotically active tissues, where they can alter DNA methylation and gene expression (37, 53). In these tissues it is essential to protect genome stability by repression of TEs so that gametes and developing organs are not harmed (54, 55).

Grafting is routinely used in agriculture to combine rootstocks and shoots (scions) with desirable characteristics, such as for grapevine, apples and tomatoes (56, 57). We have demonstrated that mobile sRNAs regulate patterns of DNA methylation genome wide, and that expression of specific genes exhibits mobile regulation. Base-resolution methylomes are being actively generated in multiple species, including crops and non-model plants (38, 58-60). Clear diversity in the epigenomes of closely related subspecies and accessions has been observed from these data (38, 58). With our demonstration that site-specific transmission of epiallele states from one

accession to another can be achieved by grafting and by *de novo* methylation of unmethylated DNA it is likely that at least some effects of grafting are due to the movement of RNA. Our findings also indicate that DNA methylomes may provide a potential new resource to growers who employ grafting. They could consider potential modification of gene expression patterns in sink tissues *via* sRNA transmission from source tissues. The mobile sRNAs can alter DNA methylation in germ line tissue (37) so that DNA methylation patterns altered by grafting may be heritable by progeny plants (37, 61).

In summary, we have shown that transmission of sRNAs from shoots to roots of *Arabidopsis* regulates DNA methylation at thousands of sites genome wide. Mobile sRNA-dependent methylation is predominantly in the non-CG context (CHG and CHH), and is associated with short Type 1 retroelements found in gene rich regions of the genome. We confirm that deposition of mobile sRNA-dependent methylation is dependent upon the *DRM1* and *DRM2* RdDM pathway, and largely independent of the *CMT2* and *CMT3* methylation pathway. Our conclusions underpin future research into why plants possess a system for communicating methylation status from shoot to root tissues.

Materials and Methods

See also Supplemental Materials and Methods.

Plant Materials and Grafting

Plants of *Arabidopsis thaliana* accessions Col-0 and C24 were used in our experiments, as well as a previously described *dcl2-1/dcl3-1/dcl4-2* triple mutant (Col-0 background) (23). The C24 plants had a GFP transgene silenced by a partial GFP inverted-repeat, termed GxGF-IR, as previously described (36). *Arabidopsis* plants were grown under 10 h supplemental fluorescent lighting at 20ºC, on vertical plates of 0.8% agar, ½ Murashige and Skoog media, pH 5.7. Micrografting was conducted seven days after germination as previously described (57, 62). Plant tissue was harvested five weeks after grafting, taking care to separate root from shoot and exclude tissue 0.3-0.5 cm surrounding the graft junction. Two independent bioreplicates were conducted for all RNA-seq samples, and two to four for each MethylC-seq sample (SI Appendix Table S3).

MethylC-seq and RNA-seq Library Construction

DNA for the MethylC-seq libraries was prepared from 3-4 pooled plants per sample using the Gentra Puregene Tissue Kit (Qiagen) according to manufacturer's instructions. MethylC-seq libraries were constructed as described previously (63). RNA-seq libraries were prepared from total RNA extracted as previously described (36). RNA-seq libraries were generated with the ScriptSeq RNA-seq Library Preparation Kit (Epicentre) according to the manufacturer's instructions.

Sequencing operations

MethylC-seq and RNA-seq libraries were sequenced using an Illumina HiSeq 2000 and 2500 platforms respectively according to the manufacturer's instructions. MethylC-seq libraries were sequenced for 101 cycles and RNA-seq libraries for 100 cycles. Image analyses and basecalling were carried out using the manufacturer supplied software and standard parameters.

Sequencing analyses

See Supplemental Materials and Methods for detailed description of analytical methods and software. In brief, sRNA data were aligned to the TAIR9 reference genome requiring perfect

matching. When analyzing MethylC-seq data we used YAMA (Yet Another Methylome Aligner), based upon Bowtie 2, to map reads to the genome (source code available at https://github.com/tjh48/YAMA). Mapping statistics of libraries are given in SI Appendix Table S3. Methylation loci were then identified independently in each context (CG, CHG, CHG) using the segmentSeq R package (64), accounting for nonconversion rates in each sample (65). Subsequently, loci exhibiting various models of differential methylation were identified using the empirical Bayesian methods for analysis of methylation data (64) with the baySeq R package (66, 67). Average methylation profiles were calculated by dividing defined regions into an equal number of windows and calculating methylation across these windows. Gene annotations were retrieved from the *Arabidopsis* Information Portal (https://www.araport.org) (68) and transposon annotations from The *Arabidopsis* Information Resource (69). Estimations of the numbers of loci expected to fit a given model were calculated using the sum of the posterior likelihoods for that model and confidence intervals were calculated by simulating ten thousand sets of true and false positives based on the posterior likelihoods. Unambiguous lists of specific differentially represented loci were identified by applying a false discovery rate of lower than 0.05 to the model, which in general represents a more conservative analysis. Associations between DMRs, genome features and sRNAs were assessed using a block-bootstrap method (70). Intersections between loci and genome features were normalized for abundance of loci and features, the sizes of loci and features and how loci and features might cluster within the genome when plotting. We did so by calculating the number of features overlapped by the DNA methylation/sRNA loci per megabase total feature length per megabase total locus length. This approach gave a visually accurate representation of significance of association. Independent measurements of statistical significance based upon the block-bootstrap analyses are presented beneath the bars also. Promoters were defined as 2000 nt preceding a TSS. Data were deposited in the European Nucleotide Archive (accession E-MTAB-3473)

Q-RT-PCR

Transcript abundance was assessed in roots of independent graft plants according as described previously (37). Fold changes in abundance of test transcripts relative to the *ACTIN 2* transcript (AT3G18780), a stable control (71), were calculated using delta delta C_T methodology (72, 73). Primer sequences (5'-3') were; GGATATTGTAGCAAGCCACAG (*ROS1* forward), CAAAGTTCATCTGGAGAAAGAC (*ROS1* reverse), GCCATCCAAGCTGTTCTCTC (*ACTIN 2* forward), CCCTCGTAGATTGGCACAGT (*ACTIN 2* reverse).

Acknowledgements

We thank Matthew D. Schultz and Yupeng He for assistance with analyses, Huaming Chen for assistance with web browser development, and Roberto Solano, Robert J. Schmitz, Taiji Kawakatsu, Chongyuan Luo, Ryan Lister, Ronan C. O'Malley and Lindsay Robinson for helpful discussions. MGL was funded by an EU Marie Curie FP7 International Outgoing Fellowship (#252475). Work in JRE's laboratory is funded by the Gordon and Betty Moore Foundation (GBMF 3034), and the National Science Foundation (MCB-1024999). JRE is an investigator of the Howard Hughes Medical Institute. CWM was funded by a Clare College (Cambridge, UK) Junior Research Fellowship. Work in the Baulcombe laboratory was supported by The Gatsby Charitable Foundation, the EU FP7 Collaborative Project grant AENEAS, and the ERC Advanced Investigator grant ERC-2013-AdG 340642. DCB is the Royal Society Edward Penley Abraham Research Professor.

References

1. Martínez de Alba AE, Elvira-Matelot E, & Vaucheret H (2013) Gene silencing in plants: a diversity of pathways. *Biochim. Biophys. Acta* 1829(12):1300-1308.
2. Hamilton A, Voinnet O, Chappell L, & Baulcombe D (2002) Two classes of short interfering RNA in RNA silencing. *EMBO J* 21(17):4671-4679.

3. Hamilton AJ & Baulcombe DC (1999) A Species of Small Antisense RNA in Posttranscriptional Gene Silencing in Plants. *Science* 286(5441):950-952.
4. Baumberger N & Baulcombe DC (2005) Arabidopsis ARGONAUTE1 is an RNA Slicer that selectively recruits microRNAs and short interfering RNAs. *Proc Natl Acad Sci U S A* 102(33):11928-11933.
5. Qi Y*, et al.* (2006) Distinct catalytic and non-catalytic roles of ARGONAUTE4 in RNA-directed DNA methylation. *Nature* 443(7114):1008-1012.
6. Nuthikattu S*, et al.* (2013) The initiation of epigenetic silencing of active transposable elements is triggered by RDR6 and 21-22 nucleotide small interfering RNAs. *Plant Phys* 162(1):116-131.
7. Bond DM & Baulcombe DC (2015) Epigenetic transitions leading to heritable, RNA-mediated de novo silencing in Arabidopsis thaliana. *Proc Natl Acad Sci U S A* 112(3):917-922.
8. Kim MY & Zilberman D (2014) DNA methylation as a system of plant genomic immunity. *Trends Plant Sci* 19(5):320-326.
9. Bernstein BE, Meissner A, & Lander ES (2007) The mammalian epigenome. *Cell* 128(4):669-681.
10. Henderson IR & Jacobsen SE (2007) Epigenetic inheritance in plants. *Nature* 447(7143):418-424.
11. Lister R*, et al.* (2008) Highly integrated single-base resolution maps of the epigenome in Arabidopsis. *Cell* 133(3):523-536.
12. Chan SW-L-L*, et al.* (2004) RNA silencing genes control de novo DNA methylation. *Science* 303(5662):1336.
13. Dowen RH*, et al.* (2012) Widespread dynamic DNA methylation in response to biotic stress. *Proc Natl Acad Sci U S A* 109(32):E2183-2191.
14. Zhong S*, et al.* (2013) Single-base resolution methylomes of tomato fruit development reveal epigenome modifications associated with ripening. *Nat Biotechnol* 31(2):154-159.
15. Gent JI*, et al.* (2013) CHH islands: de novo DNA methylation in near-gene chromatin regulation in maize. *Genome Res* 23(4):628-637.
16. Zilberman D*, et al.* (2004) Role of Arabidopsis ARGONAUTE4 in RNA-directed DNA methylation triggered by inverted repeats. *Curr Biol* 14(13):1214-1220.
17. Zilberman D, Cao X, & Jacobsen SE (2003) ARGONAUTE4 control of locus-specific siRNA accumulation and DNA and histone methylation. *Science* 299(5607):716-719.
18. Schmitz RJ*, et al.* (2011) Transgenerational epigenetic instability is a source of novel methylation variants. *Science* 334(6054):369-373.
19. Becker C & Weigel D (2012) Epigenetic variation: origin and transgenerational inheritance. *Curr Opin Plant Biol* 15(5):562-567.
20. Richards EJ (2006) Inherited epigenetic variation--revisiting soft inheritance. *Nat Rev Gen* 7(5):395-401.
21. Papp I*, et al.* (2003) Evidence for Nuclear Processing of Plant Micro RNA and Short Interfering RNA Precursors. *Plant Phys* 132(3):1382-1390.
22. Wierzbicki AT, Ream TS, Haag JR, & Pikaard CS (2009) RNA polymerase V transcription guides ARGONAUTE4 to chromatin. *Nat Gen* 41(5):630-634.
23. Henderson IR*, et al.* (2006) Dissecting Arabidopsis thaliana DICER function in small RNA processing, gene silencing and DNA methylation patterning. *Nat Gen* 38(6):721-725.
24. Mosher RA, Schwach F, Studholme D, & Baulcombe DC (2008) PolIVb influences RNA-directed DNA methylation independently of its role in siRNA biogenesis. *Proc Natl Acad Sci U S A* 105(8):3145-3150.
25. Zhong X*, et al.* (2014) Molecular mechanism of action of plant DRM de novo DNA methyltransferases. *Cell* 157(5):1050-1060.
26. Woo HR, Dittmer TA, & Richards EJ (2008) Three SRA-Domain Methylcytosine-Binding Proteins Cooperate to Maintain Global CpG Methylation and Epigenetic Silencing in Arabidopsis. *PLoS Gen* 4(8):e1000156.
27. Chan SW-L, Henderson IR, & Jacobsen SE (2005) Gardening the genome: DNA methylation in Arabidopsis thaliana. *Nat Rev Gen* 6(5):351-360.

28. Kim J, Kim JH, Richards EJ, Chung KM, & Woo HR (2014) Arabidopsis VIM proteins regulate epigenetic silencing by modulating DNA methylation and histone modification in cooperation with MET1. *Mol Plant* 7(9):1470-1485.
29. Law JA & Jacobsen SE (2010) Establishing, maintaining and modifying DNA methylation patterns in plants and animals. *Nat Rev Gen* 11(3):204-220.
30. Du J*, et al.* (2012) Dual binding of chromomethylase domains to H3K9me2-containing nucleosomes directs DNA methylation in plants. *Cell* 151(1):167-180.
31. Stroud H, Greenberg MVC, Feng S, Bernatavichute YV, & Jacobsen SE (2013) Comprehensive analysis of silencing mutants reveals complex regulation of the Arabidopsis methylome. *Cell* 152(1-2):352-364.
32. Johnson LM*, et al.* (2014) SRA- and SET-domain-containing proteins link RNA polymerase V occupancy to DNA methylation. *Nature*.
33. Haag JR & Pikaard CS (2011) Multisubunit RNA polymerases IV and V: purveyors of non-coding RNA for plant gene silencing. *Nat Rev Mol Cell Biol* 12(8):483-492.
34. Stroud H*, et al.* (2014) Non-CG methylation patterns shape the epigenetic landscape in Arabidopsis. *Nat Struct Mol Biol* 21(1):64-72.
35. Melnyk CW, Molnar A, & Baulcombe DC (2011) Intercellular and systemic movement of RNA silencing signals. *EMBO J* 30(17):3553-3563.
36. Molnar A*, et al.* (2010) Small silencing RNAs in plants are mobile and direct epigenetic modification in recipient cells. *Science* 328(5980):872-875.
37. Melnyk CW, Molnar A, Bassett A, & Baulcombe DC (2011) Mobile 24 nt small RNAs direct transcriptional gene silencing in the root meristems of Arabidopsis thaliana. *Curr Biol* 21(19):1678-1683.
38. Schmitz RJ*, et al.* (2013) Patterns of population epigenomic diversity. *Nature*.
39. Laubinger S*, et al.* (2010) Global effects of the small RNA biogenesis machinery on the Arabidopsis thaliana transcriptome. *Proc Natl Acad Sci U S A* 107(41):17466-17473.
40. Bond DM & Baulcombe DC (2014) Small RNAs and heritable epigenetic variation in plants. *Trends Cell Biol* 24(2):100-107.
41. Groszmann M*, et al.* (2011) Changes in 24-nt siRNA levels in Arabidopsis hybrids suggest an epigenetic contribution to hybrid vigor. *Proc Natl Acad Sci U S A* 108(6):2617-2622.
42. McCormick KP, Willmann MR, & Meyers BC (2011) Experimental design, preprocessing, normalization and differential expression analysis of small RNA sequencing experiments. *Silence* 2(1):2.
43. Dunoyer P*, et al.* (2010) An endogenous, systemic RNAi pathway in plants. *EMBO J* 29(10):1699-1712.
44. Fedoroff NV (2012) Presidential address. Transposable elements, epigenetics, and genome evolution. *Science* 338(6108):758-767.
45. Howell MD*, et al.* (2007) Genome-wide analysis of the RNA-DEPENDENT RNA POLYMERASE6/DICER-LIKE4 pathway in Arabidopsis reveals dependency on miRNA- and tasiRNA-directed targeting. *Plant Cell* 19(3):926-942.
46. Gasciolli V, Mallory AC, Bartel DP, & Vaucheret Hae (2005) Partially redundant functions of Arabidopsis DICER-like enzymes and a role for DCL4 in producing trans-acting siRNAs. *Curr Biol* 15(16):1494-1500.
47. Dunoyer P, Himber C, & Voinnet O (2005) DICER-LIKE 4 is required for RNA interference and produces the 21-nucleotide small interfering RNA component of the plant cell-to-cell silencing signal. *Nat Gen* 37(12):1356-1360.
48. Golden TA*, et al.* (2002) SHORT INTEGUMENTS1/SUSPENSOR1/CARPEL FACTORY, a Dicer homolog, is a maternal effect gene required for embryo development in Arabidopsis. *Plant Phys* 130(2):808-822.
49. Lenoir A*, et al.* (2001) The evolutionary origin and genomic organization of SINEs in Arabidopsis thaliana. *Mol. Biol. Evol.* 18(12):2315-2322.
50. Wang X, Weigel D, & Smith LM (2013) Transposon variants and their effects on gene expression in Arabidopsis. *PLoS Gen* 9(2):e1003255.

51. de la Chaux N, Tsuchimatsu T, Shimizu KK, & Wagner A (2012) The predominantly selfing plant Arabidopsis thaliana experienced a recent reduction in transposable element abundance compared to its outcrossing relative Arabidopsis lyrata. *Mob DNA* 3(1):2.
52. Zhang X & Wessler SR (2004) Genome-wide comparative analysis of the transposable elements in the related species Arabidopsis thaliana and Brassica oleracea. *Proc Natl Acad Sci U S A* 101(15):5589-5594.
53. Zhang W*, et al.* (2014) Graft-transmissible Movement of Inverted Repeat - induced siRNA Signals into Flowers. *Plant J* 80(1):106-21.
54. Ibarra CA*, et al.* (2012) Active DNA demethylation in plant companion cells reinforces transposon methylation in gametes. *Science* 337(6100):1360-1364.
55. Hsieh TF*, et al.* (2009) Genome-wide demethylation of Arabidopsis endosperm. *Science* 324(5933):1451-1454.
56. Melnyk CW & Meyerowitz EM (2015) Plant grafting. *Curr Biol* 25(5):R183-188.
57. Melnyk Charles W, Schuster C, Leyser O, & Meyerowitz Elliot M (2015) A Developmental Framework for Graft Formation and Vascular Reconnection in Arabidopsis thaliana. *Curr Biol* 25(10):1306-1318.
58. Li X*, et al.* (2012) Single-base resolution maps of cultivated and wild rice methylomes and regulatory roles of DNA methylation in plant gene expression. *BMC Genomics* 13:300.
59. Schmitz RJ*, et al.* (2013) Epigenome-wide inheritance of cytosine methylation variants in a recombinant inbred population. *Genome Res* 23(10):1663-1674.
60. Takuno S & Gaut BS (2013) Gene body methylation is conserved between plant orthologs and is of evolutionary consequence. *Proc Natl Acad Sci U S A* 110(5):1797-1802.
61. Wu R*, et al.* (2013) Inter-species grafting caused extensive and heritable alterations of DNA methylation in Solanaceae plants. *PloS one* 8(4):e61995.
62. Turnbull CGN, Booker JP, & Leyser HMO (2002) Micrografting techniques for testing long-distance signalling in Arabidopsis. *Plant J* 32(2):255-262.
63. Urich MA, Nery JR, Lister R, Schmitz RJ, & Ecker JR (2015) MethylC-seq library preparation for base-resolution whole-genome bisulfite sequencing. *Nature Protoc* 10(3):475-483.
64. Hardcastle TJ (2015) Discovery of methylation loci and analyses of differential methylation from replicated high-throughput sequencing data. *bioRxiv*.
65. Team RDC (2008) R: A language and environment for statistical computing. R Foundation for Statistical Computing.
66. Hardcastle TJ & Kelly KA (2013) Empirical Bayesian analysis of paired high-throughput sequencing data with a beta-binomial distribution. *BMC Bioinformatics* 14:135.
67. Hardcastle TJ (2015) Generalised empirical Bayesian methods for discovery of differential data in high-throughput biology. *Bioinformatics*.
68. Consortium IAI (2012) Taking the next step: building an Arabidopsis information portal. *Plant Cell* 24(6):2248-2256.
69. Huala E*, et al.* (2001) The Arabidopsis Information Resource (TAIR): a comprehensive database and web-based information retrieval, analysis, and visualization system for a model plant. *Nucleic acids research* 29(1):102-105.
70. Bickel PJ & Freedman DA (1981) Some Asymptotic Theory for the Bootstrap. *Ann. Statist.* 9(6):1196-1217.
71. Havecker ER, Wallbridge LM, Fedito P, Hardcastle TJ, & Baulcombe DC (2012) Metastable differentially methylated regions within Arabidopsis inbred populations are associated with modified expression of non-coding transcripts. *PloS One* 7(9):e45242.
72. Yuan JS, Wang D, & Stewart CN (2008) Statistical methods for efficiency adjusted real-time PCR quantification. *Biotechnol J* 3(1):112-123.
73. Livak KJ & Schmittgen TD (2001) Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods* 25(4):402-408.

Figure Legends

Figure 1. Genomic loci in *Arabidopsis* roots were classified according to their combination of sRNA and DNA methylation status. Grafts were made between shoots and roots of various *Arabidopsis* genotypes (combinations indicated at top, denoted shoot/root). Target loci of sRNAs and DNA methylation were identified by analyzing MethylC-seq and sRNA-seq data from roots of all graft combinations. Each locus classification was defined by a specific combination of sRNA and DNA methylation (mC) levels across the five graft combinations. "+" denotes a relatively high level of mC or sRNA, whilst "-" denotes a relatively low level. Classifications were designated A-C and are indicated on the left. See also SI Appendix Fig. S1.

Figure 2. DNA methylation of mobile sRNA associated loci (class A, direct, and B, indirect) is reduced most strongly by *drm1/2* mutation. The plots show the mean proportion of DNA methylation across class A and B loci in mutants of RNA-directed DNA methylation (RdDM) pathway components and wild-type (WT) plants. The legend on the right indicates which color represents each mutant. Data are shown separately for non-CG (CHG and CHH) sequence contexts, which had the strongest association with mobile sRNAs. The proportion of methylation was calculated per base (between 0 and 1, unmethylated to fully methylated) for all loci within a class using published methylC-seq data from RdDM mutants (31). Loci were then normalized to same size and the mean proportion of DNA methylation calculated across them. The profiles of the mean proportion of methylation across the size-normalized loci are plotted between the dashed vertical lines (indicated by the solid black bar labeled loci). Mean proportion of methylation in flanking DNA, 4kb upstream and downstream of the loci, are indicated to the left and right of the dashed vertical lines, respectively. Total number of loci assessed is given in parentheses at the top right corner of each plot.

Figure 3. Mobile sRNA directly targets DNA methylation at promoters containing transposable elements. (A) CHG and CHH (non-CG) context class A loci are enriched in TEs and in promoters containing TEs, but are depleted in genes and coding sequences (CDS). Very few CG context loci were identified. (B) Non-CG context class A loci preferentially target classes of Type 1 retroelement, indicated by significant enrichment. They also, less strongly, target some Type 2 DNA element classes. (C) The class A loci are targeted by graft-transmissible 23-24 nt sRNAs (the known mobile class). Significance levels; # = $0 < p < 10^{-5}$, + = $10^{-4} < p < 10^{-5}$, * = $10^{-3} < p < 10^{-4}$, ! = $10^{-2} < p < 10^{-3}$, blue = under-represented, red = over-represented (relative to background). (A) CDS – coding sequence, TE – transposable element, UTR – untranslated region. (A, B) Y-axis units normalize the feature/locus overlap by both sum of genome feature size and sum of methylation locus size, which permits comparison between columns. These units are number of annotated features per total megabase (MB) of named feature per total MB of methylation in locus class. Numbers of loci in each DNA methylation context shown within parentheses in the legend. (C) Loci were normalized to same size on graph, indicated by dark bar on position axis, flanked by sRNA coverage 4kb upstream and downstream of locus. Numbers of loci assessed are given in parentheses.

Figure 4. Loci where DNA methylation is indirectly targeted by mobile sRNAs (class B) exhibit similar characteristics to loci that are directly targeted (class A). (A) Non-CG (CHG and CHH) context B loci are significantly enriched in TEs and in promoters containing TEs, but are depleted in promoters that do not contain TEs, genes and coding sequences (CDS). Very few CG context loci were identified. (B) Non-CG context B loci target most classes of TEs, but show most significant association with classes of Type 1 retroelement. They also, less strongly, target some Type 2 DNA element classes. Y-axis units normalize the feature/locus overlap by both sum of feature size and sum of methylation locus size, which permits comparison between columns. These units are number of annotated features per total megabase (MB) of named feature per total MB of methylation in locus class. Significance levels; # = $0 < p < 10^{-5}$, + = $10^{-4} < p < 10^{-5}$, * = $10^{-3} < p < 10^{-4}$, ! = $10^{-2} < p < 10^{-3}$, blue = under-represented, red = over-represented (relative to background). TE – transposable element, UTR – untranslated region. Numbers of loci assessed are given in parentheses.

Figure 5. C24-derived sRNAs can target DNA methylation *de novo* at unmethylated regions of the Col-0 genome. Three such *de novo* loci are shown in AnnoJ genome browser screenshots. The screenshots display sRNA reads and methylated cytosine residues (mC) for graft combinations C24/C24, Col/Col and C24/Col. Results from two independent biological replicates, suffixed 1 and 2, are shown for each graft combination. Note the groups of sRNA reads present only in C24/C24 and C24/Col, which correlate with the presence of mC. Red and green sRNAs indicate that they map to the Watson (W) and Crick (C) strands, respectively. Gold, blue and pink mC positions indicate methylation in the CG, CHG and CHH context, respectively.